

Reuse Working Group – Report Back

Co-chairs:

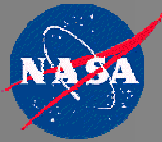
Robert Wolfe, NASA GSFC

Al Fleig, PITA Analytic Sciences

7th Earth Science Data Systems Working Group Meeting

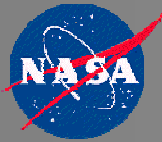
Philadelphia, PA

October 21–23, 2008



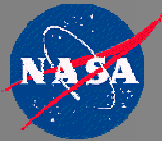
Reuse Agenda and Summary

-
- Discussion of “reuse” and “reusability” by Michael Leyton
 - Developed improved definitions of these terms
 - Discussion of decadal survey missions and reuse
 - Service Oriented Architecture and other new technologies (added)
 - *Invited speaker: Bill Frakes, Virginia Tech*
 - Discussion of data life cycle issues
 - *Invited speaker: Eduardo Almeida, Recife Center for Advanced Studies and Systems (CESAR)*
 - Discussion of Reuse Readiness Levels
 - Planning for 2009



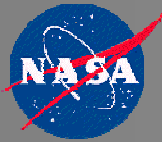
Michael Leyton proposed the following statements, which the WG accepted:

- The reuse of a software artifact is its integration into another context.
- Reusability (reuse readiness) is the extent to which a software artifact is integratable into another context
- The purpose of reuse is to reduce cost, time, effort, and risk; and to increase productivity, quality, performance, and interoperability.



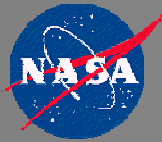
If we were to do EOS over again, what would we have done differently?

- Implement a mission and algorithm software (and documentation) repository
 - Current repository is distributed (STs, Data Centers, SIPS, etc.)
 - A set of best practices for reuse should be developed (and then evolve these practices into standards and policies) – Is there a best practices document for software reuse? – See “reuse guideline” documents on bottom up reuse (on Reuse portal).
 - Science community needs to be on the same wavelength – needs to involve data systems folks in the discussion of future planning for data system policies
- Don't be so ambitious – too many promises about great leaps forward
- Spend more of overall mission budget on data system development (not just paper) – algorithm development will cost more and take longer than expected
- Overall structure should be less monolithic – it started that way but became more distributed (i.e. SIPS) – but this may have had a detrimental impact on preserving algorithms and software – the requirements for the SIPS should include requirements for preserving (and delivering) algorithms and software
- It will be a few years after launch before good quality data is produced and that algorithm (and software) revisions will be needed for the life of the mission
- Algorithm documentation needs to be updated before/after major changes to the software
- Standards and protocols are needed for preserving algorithms and software in a distributed system
- **Future reuse (persistent use) is important**



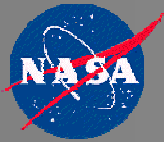
What role should Service Oriented Architecture (SOA) and other new technologies play?

- Cloud computing should be considered as an option – but data ownership and control is a key issue – there are large potential cost savings – but we need aware other risks – service level agreements (SLAs) – Cloud computing could be done within (internal to) NASA
- SOA – if done right, it will have a large cost savings in the future – but raises new issues (risks) – there may be provenance issues – trust but verify (often) – science testing in a known environment (to verify SOA results) is needed to produce good quality products
- Verification of correct results from SOA is different and more difficult – many more independent changes are possible – different approach to verification may be needed (self checking)
- Reproducibility could be a problem – how do you later reproduce a production run after all of the versions of the intermediate steps have changed? This could be handled by “snap shots” of end-to-end change at specific points in the process.
- Preserving the software (for persistent use or future reuse) is more difficult.

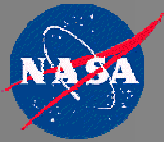


What should be done to prepare for the decadal survey missions?

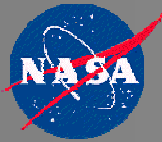
- Input data from outside organizations being used in the algorithm needs to be preserved (either by outside organization or internally) – this is also an issue for current systems – reproducibility issue
- RES(s) – either central or distributed should be established – but links are needed between distributed RESs
- Linkages between data and software to produce data (provenance) should be established and maintained
- Open Source software should be used whenever feasible – because of the risks of proprietary software – this helps with preservation of software
- Policies for use of web services may need to be established – wrt. software and algorithm capture and preservation
- New process for prioritization of algorithms and data sets is needed (community vetted)
- What are the high leverage points (in terms cost and schedule) for software reuse in the decadal survey missions? Studies are needed to figure this out.



- Future Reuse
 - RRL for tailored future use; extensibility is not as important, no binaries, etc.
 - How does this relate to the OAIS (Open Archive Information System) model – could a tailored RRL be added to the OAIS model?
- Reuse of other data life cycle models
 - To be reused by new DLC WG
- Identify and flesh out use cases for future reuse
 - 4 or 5 use cases
- RES – is important for future reuse
 - Future reuse should be used to help leverage implementation of a RES
 - Should informed by with OAIS standard for submission (archive) information package
 - We could help evolve the OAIS standard wrt. software preservation and future reuse; better representation information (information for rendering or making data useable, e.g. documentation)

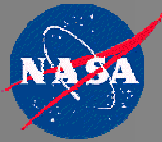


- Policy issues for Future Reuse (Persistent Use)
 - Proprietary software needs to be available for future use
 - Escrow approach needed to store software until proprietary period expires
 - ITAR (International Traffic in Arms Regulations) – similar to proprietary issue; restricted distribution
 - Others?
- Establish and maintain relationships between software and data
 - Version of software needs to be tied to data produced (or read) by software
- Making data (re)usable in the future – is this in scope?
 - What are the best practices for producing reusable data?
 - What should the measures projects do for reuse?



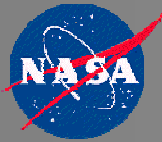
Draft Reuse Readiness Levels

- Did a review of the Reuse Readiness Level (RRL) topic area levels
 - More work is still needed
- Next steps:
 - Produce a full document – background, justification, etc.
 - Develop use cases for developer and consumer
 - Use the RRLs to assess real software (try them out)
 - Use feedback to revise RRLs
 - Have face-to-face meeting in April to finalize levels

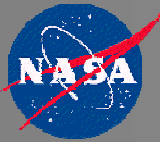


2008 NASA ESDS Software Reuse Award Recipients (alphabetically by project name or individual's last name):

- Contribution Award Category: 2 recipients
 - [Mercury Consortium](#) (Oak Ridge National Laboratory)
Mercury Distributed Metadata Management, Data Discovery and Access System
 - [UAHuntsville Subset Team](#) (University of Alabama at Huntsville)
HDF-EOS Web-based subsetter (HEW) family of subsetting software
- Utilization Award Category: 1 recipient
 - [Data Management Systems and Technologies Group](#) (NASA JPL)
Object-Oriented Data Technology (OODT) Catalog and Archive Service (CAS)
- Peer Education Award Category: 2 recipients
 - [Dr. Victor E. Delnore](#) (NASA LaRC)
General peer-education; ESDS Software Reuse WG activities
 - [Dr. Robert R. Downs](#) (Columbia University)
General peer-education; ESDS Software Reuse WG activities



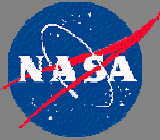
- Reuse Enablement System (RES)
 - ***Develop implementation plan – need plan for deployment***
 - Continue developing and testing prototype – initial prototype is done, some small changes in work; test with a number of software artefacts
 - ***Vet RES policies (internal and external)***
 - ***Deploy the prototype for NASA internal use***
- Reuse portal web site
 - Provide more content and keep up to date
 - Promote portal to community
 - Add RES roadmap (and schedule for RES deployment eventually)



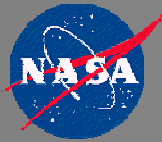
Plan for 2009 – Tasks (2 of 3)

- Provide incentives for reuse
 - **Peer recognition process** – continue with peer-award next year;
continue to work to develop a NASA reuse award process
 - Work on recommendation/justification for HQ to develop a funding opportunity to make assets/components reusable (within the ES community)
- Metrics/measurements
 - Continue to generate/analyze statistics for portal web site
 - **Assess effort required to package assets for reuse**
 - Also quantify benefits of open source release of assets
 - Still need work on small components, small vs. big process – are they different for tech. transfer?
 - RES could be used to collect metrics – ask providers how much reuse they did to produce the artefact; ask the consumers how much saving they achieved by reusing the artefact
 - Develop impact metrics from Peer-awardees
 - Summarize the NPP Reuse study up on portal – as an impact metric

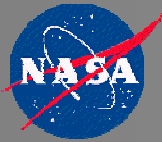
Higher priority



- Promote reuse
 - Continue publications in journals and presentations at conferences
 - Prototype a process for facilitating reuse through mentoring – including IP issues
 - ***Continue developing reuse readiness level scale → work with Standards WG***
- Policy
 - ***Continue working with Office of Technology Transfer to facilitate software release process → work to understand and change the process – lower the barrier(s) for certain types of S/W***
- Data Life Cycle (DLC)
 - ***Work with “new” DLC group to help with persistent (re)use areas***
- Decadal survey
 - ***Tie reuse efforts to the decadal survey missions***
 - ***Prepare for reuse as key part of (and asset for) new missions***



-
- Reaching out to other ES domains (e.g., modeling community – ESMF, National Forum for Geoscience Information Technology – FGIT, NOAA, EPA, ESIP, etc.)
 - Follow-up on establishing a ESIP reuse sub-group



-
- Software Reuse Portal Web Site
 - <http://www.esdswg.com/softwarereuse>
 - Collaboration Web Site
 - <http://www.sciencedatasystems.org/reuse/default.aspx>
 - Mailing List
 - Send e-mail to Support Contact (below)
 - Monthly Telecons
 - Third Wednesday of the month @ 2 pm Eastern time
 - Support Contact
 - Jim Marshall (James.J.Marshall@nasa.gov)